AUTOMATIC DIAGNOSIS OF FETAL HEART RATE: COMPARISON OF DIFFERENT METHODOLOGICAL APPROACHES

G. Magenes[1], M.G. Signorini[2], R.Sassi[2]
[1]Dipartimento di Informatica e Sistemistica, Unversity of Pavia, Italy
[2]Dipartimento di Bioingegneria, Politecnico di Milano, Italy

Abstract- **The cardiotocograpy (CTG) is the clinical, traditional, noninvasive approach to monitor the fetal condition antepartum. CTG analysis is focused on the detection of fetal heart rate parameters from which the clinicians can identify by eye inspection some patterns associated to fetal activity. However this qualitative method rarely can detect the emergence of fetal pathologies. This study aims at finding new algorithms which can enhance the differences among the normal CTG signals and those presenting anomalies due to a pathological status.**
On a database of more than 500 recordings, we tested different classification methods to identify normals from potential pathological fetuses. A Multilayer Perceptron (MLP) neural network and an Adaptive Neuro-Fuzzy Inference System (ANFIS) were compared with classical statistical methods. Both the neural and neuro-fuzzy approaches seem to give better results than any tested statistical classifier.
***Keywords* - Fetal Heart Rate, Neural Networks, Fuzzy Systems, Multivariate Methods**

I. INTRODUCTION

The cardiotocography (CTG) is regularly monitored in the clinical routine antepartum and during the labour in order to prevent a possible fetal sufferance status. It consists of the simultaneous recording and printout of two signals: the heartbeat frequency of the fetus and the toco signal, relative to the uterine contractions.
It was only at the end of the 60's, when the fetal heartbeat could be rather easily detected by means of ultrasound (the Doppler-shift) or through the application of direct electrocardiography, that cardiotocography became popular as the method to monitor the condition of the fetus. This modality provides not only continuous heart rate information, but also fetal heart rate changes in response to uterine contractions. Currently the majority of obstetric decisions to assist delivery of the baby by artificial means (caesarean section, forceps or vacuum extraction) for reasons of suspected fetal distress, relies on information gathered through the application of cardiotocography. It is the obstetrician's reassurance that if the fetal heart rate (FHR) pattern is normal then there is the nearly 100% certainty that the fetus is in a good condition, which has made cardiotocography so attractive and has induced its widespread use [1].
The classical cardiotocographic analysis by simple eye inspection has drastically reduced the incidence of deaths during the labour and also in premature newborns, although the presence of many false positives. This problem can be attributed to the inter and intra-observer variability, in the interpretation of the CTG signals, due in the first case to the different levels of experience of the various specialists , and in the second to suggestive factors (stress, environmental conditions). The reading of the printout in most cases is mainly a subjective process and, in certain cases, may lead to a wrong decision. The obvious consequence, in a false positive case, can be to decide to execute a caesarean when it is not necessary, or , in a false negative case to let the pregnancy go on, when a caesarean should have been done.
Although a number methods for judging CTG recordings were proposed [2, 3, 4] and few systems able of automatically computing quantitative parameters were developed [5,6], no one of them showed a strong reliability in predicting the fetal well-being
The aim of the present work is to test and compare several classification techniques of the CTG signal, allowing to evolve into a reliable automatic system of "reading" and analysing the CTG tracings in the hope of diagnosing any eventual fetal suffering status.

II. METHODOLOGY

A. Data collection

The data were recorded during two years in a University Clinic in Rome, Italy. 815 CTG recordings were collected from four identical devices (HP M135XA). For 549 of them we even knew the diagnosis of the physician at delivery (weight, type of delivery, Apgar score). Each recording lasted at least 30 minutes and it contained both the cardiographic series and the toco trace. We focused on four potential pathological states: (i) nutrition alterations caused by maternal hypertension (H),(ii) intra-uterine growth retardation (IUGR), (iii) nutrition alterations caused by maternal diabetes (DG), and (iv) fetal macrosomia (MACRO). The gestational age was in the range 28–42 weeks.

B. FHR Preprocessing

A quality index quantifies three different levels of the FHR signal (optimal (green), acceptable quality (yellow) and insufficient quality – signal unavailable (red)). The evaluation is based on the output of the autocorrelation procedure implemented in the HP1350. Signals were recorded at the highest available sampling frequency (2Hz). Each FHR series underwent a subdivision into 3-minutes segments (360 points) after removing the red-quality points at the beginning of the sequence. We obtained a set of 549 recordings, in which we further considered only those with 5 segments (360 points each) of sufficient quality at least, discarding the other ones. This second level of refinement led us to a subset of 362 valid recordings which are summarized in Table 1.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Automatic Diagnosis of Fetal Heart Rate: Comparison of Different Methodological Approaches	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Dipartimento di Informatica e Sistemistica, University of Pavia, Italy	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 803 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom.	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 4	

TABLE I
ANALYZED SUBJECTS

ID	Pathol. State	N° Patients	N°Recordings
1	N	154	200
2	H	32	53
3	IUGR	23	40
4	DG	19	38
5	MACRO	24	31
Total		**252**	**362**

C. Parameters Extraction

In order to extract the diagnostic information from the CTG signals, we calculated a series of parameters in time and in frequency domain. They are summarized in Table 1. Most of them can be related to the physiological mechanisms that perform the control of the HR signal. The final goal of our research was to investigate if a group of indexes, ***x***, is able to characterize the signal, that is, if it is possible to automatically allocate, by means of a classification technique, a fetus to a pathological state according to the value of ***x***.

When possible, parameters have been calculated more then once (on each sufficient quality 3-minutes-segment and on each minute) and subsequently averaged.

A set of 15 parameters plus the gestational age of the fetus, constituted the multivariate variable ***x***, which was used for the classification process.

Parameters might be grouped as:

- *Morphological* - large and small accelerations per hour, decelerations per hour and contractions per hour
- *Time domain* - FHR mean over a minute (mean FHR), FHR standard deviation (std FHR), Delta FHR, Short term variability (STV), Long term irregularity (LTI), Interval Index (II),
- *Frequency domain* from autoregressive power spectrum estimation - LF-power, MF-power, HF-power and LF/(MF+HF)) and
- *Regularity* parameters - approximate entropy (ApEn) [7]

Accelerations and decelerations were computed automatically by the Mantel's algorithm [3], as well as the uterine contractions, that resulted from the application of a modified version of the FHR baseline computation applied to the tocographic signal. The remaining parameters were computed as reported in [8].

A few standard statistical analysis were performed on the parameters set to verify the degree of linear dependence. As part of the computations involved in several methods, the covariance matrix of the variables in the model is inverted. Variables linearly dependent on the other ones would lead to ill-conditioned matrices, which can not be inverted. Moreover, completely redundant variables would only make computations more complex. From the analysis of the covariance matrix [9], the condition number resulted always acceptable.

TABLE 2
LIST OF COMPUTED PARAMETERS

computed on the whole signal:	
(1) n. large accelerat. / hour	(2) n.small accelerat. / hour
(3) n. decelerations / hour	(4) n .contractions / hour
computed on each 3-minutes SQ-segments	
(5) mean FHR (ms)	(6) std FHR (ms)
(7) LTI (ms)	(8) LF-power (ms^2)
(9) MF-power (ms^2)	(10) HF-power (ms^2)
(11) LF/(MF+HF)	(12) ApEn(1, 0.2)
computed on each minute in each 3-minutes SQ-segments	
(13) Delta (ms)	(14) STV (ms)
(15) Interval Index	

D. Classification With Multivariate Methods

The object of the multivariate statistical analysis proposed in this paper are variables which have been measured in human fetuses by means of cardiotocographic equipments.

The data set is a matrix X ($n \times p$), where n is the number of observations (362 recordings) and p the number of variables parameters computed on each recording). A single row of X may be though as an observation extracted from a multivariate distribution.

Multivariate methods can be separated in two main groups: (i) methods that assume a given structure into g groups and specify to which of them each case belongs; (ii) methods that seek for discovering a possible structure in the dataset, eventually obtaining a separation into groups [10]. Following the typical terminology of pattern recognition, the first ones are called supervised methods and the second ones unsupervised. Supervised methods try to allocate future cases (for example, future CTG recordings) to one of the g pre-specified classes in which the current observations are collected. Modern statistics refers to the process of case allocating into predefined classes (medical diagnosis, for example) as "classification" [11]. Almost all classification methods can be seen as ways to approximate an optimal classifier, the Bayes rule. Given a future case x, the classifier finds the class k with the largest posterior probability $p(k \mid x)$ and allocates the case to this class. The posterior probability are learned from a training set, a collection of examples, already classified (by experts or physicians, for example). This approach, where the estimated probabilities $p(k \mid x)$ are used as true probabilities, can result in over-fitting, by performing very well only on the training set but not on any future cases. To avoid this problem, the available data are usually split into two subsets, a training and a test set. The first one is used to estimate the classification model; the second acts as a group of future cases and is classified with the model previously obtained. In this way over-fitting is excluded (the second set was not employed when the classifier was constructed) and a reliable estimation of the performances of the classification process is achieved.

In our approach we decided to use the following statistical methods:

- Linear & Quadratic Discriminant Analysis (LDA and QDA)
- Logistic Discriminant Analysis
- K-nearest neighbour classifiers

E. Parameters Reduction

In order to use efficiently soft computing methods as neural networks and fuzzy systems, we needed to reduce the number of variables. This is mainly due to several wellknown reasons: the difficulty of managing fuzzy systems with a quite large number of inputs, the risk of overfitting with a large number of neurons in the NN with a small training set, the convergence time of learning procedures and the probability to fall in a local minimum in a hyperspace of 16 dimensions. Moreover it would be possible that a few variables were not relevant to the classification process and were acting as noise. Unfortunately, both the MLP and the ANFIS are essentially nonlinear systems and they do not allow to uniquely distinguish which parameters are less important then the other ones inside the classification process. Therefore, several different approaches were attempted. Most of them are relevant to the construction of a linear model. Nevertheless they can give interesting insight and a possible starting point in the variable selection process which must be performed by successive experiments, anyhow.
We applied Mono-variate t-test, Multi-variate F-test and Principal Component Analysis (PCA) for reducing the number of input variables. By means of these methods we extracted 5 variables which demonstrated the highest sensitivity to discrimination among normals and pathological fetuses.. They are reported in table 3.

TABLE 3
REDUCED SET OF PARAMETERS

computed on the whole signal	1.	large accelerations per hour
	2.	small accelerations per hour
computed on each 3-minutes SQ-segments	3.	LTI (ms)
	4.	LF/(MF+HF)
	5.	ApEn(1, 0.2)

III. RESULTS

As we employed several supervised techniques, a validation procedure was needed in order to test the generalization properties of the different classification methods. Because of the limited number of recordings, we decided to apply a standard crossvalidation technique. At first, 7 non overlapping subsets, of 50 recordings each, were randomly chosen from the full set of 362 exams. Then, with each supervised method, a 7-fold cross-validation technique was employed, using the same subsets partition (12 exams never entered any test set, though they were always contained in the training partition). This procedure ensures a fair comparison among different methods. The validation technique consisted of a "leave fifty out" procedure. Besides, the whole population was divided in two groups: normal (labelled "1"), if the baby at delivery was regarded as N, and pathological (labelled "2") when the fetus was included in states H, IUGR, DG and MACRO.

Multilayer Perceptron (MLP)

We tested different MLP architectures, all presenting 5 input and 1 output neurons. The internal hidden layers were composed by neurons having a tansigmoid activation function, namely

$$y = 2/[1 + \exp(-2x)] - 1$$

The output of the network was quantized in two values, with a static threshold set at zero ($-1 \equiv$ "pathological" and $1 \equiv$ "normal"). The MLPs were trained by the adaptive backpropagation method and the test was performed following the crossvalidation procedure reported above. Input CTG parameters in each training set and the corresponding actual output groups were used to train the network (30000 training epochs), until an acceptable error goal was achieved. Among the various architectures the best one resulted with three layers, composed by 12, 8 and 1 neurons, respectively. The classification performance of the NN is reported in table 4. The MLP performed better then any other technique which has been evaluated in this work, showing a 20% misclassification rate and an appreciable sensitivity and specificity , both reaching approximatively 80%.

Adaptive Neuro Fuzzy Inference System (ANFIS)

A further approach to our classification problem consisted of applying a Neuro Fuzzy inference system for discriminating among normals and pathological tracings. The classifier adopts the "Sugeno" metrics and it has been designed by means of the Matlab Fuzzy Toolbox. It receives as input the five parameters and the gestational age of the fetus and produces as ouput one of the two classes (normal or pathological). The advantage of using this methodology basically resides on the fact that while maintaining the fuzzy approach (in alternative to all previous classification methods which are "crisp"), it can be trained exactly as a supervised neural network. Both the rules and the membership functions are optimized by the learning procedure to obtain the minimum error on the input-output training set. This means that the designer is not burdened by the usual tasks of fuzzy logic which impose to write out the inference rules and to determine the membership functions.
The ANFIS model is structured to generate a number of inference of rules given by the simple relationship

$$n° \, Rules = (n° \, MF)^{n° \, INPUT}$$

where $n°MF$ is the number of levels of the memebrship functions and *INPUT* is the number of variables. In our case the *INPUT* was 6 (5+gestational age) and the only reasonable $n°MF$ was 2 ($n°$ *Rules* = 64) in order to avoid overfitting. After the crossvalidation procedure the performance of our ANFIS is summarized in table 4. An "a-posteriori" analysis of the inference rules automatically generated by the learning procedure showed that *n° Rules* can be manually reduced to 37 without deteriorating so much the global performance of the classifier (25% of misclassification).

IV. DISCUSSION

At present, automated methods have limited clinical applications in cardiotocography. A relevant amount of this unsatisfactory performance resides on the weakness of methods used for classifying fetal condition generating risk alarms during pregnancy [8]. Moreover, even if heart rate variability became an integral part in fetal evaluation, from the clinical point of view the lack of standardization makes any comparison very difficult. In the present work we tried to move a step forward towards an automated CTG risk alerts generator, that might help the physician in drawing the final diagnosis. The work was performed at two different levels.

TABLE 4
COMPARISON OF CLASSIFIERS

Input 5 parameter set (+ gestational age in ANFIS)			
	Misclas.Rate	Sensitivity	Specificity
Statistical Classifiers			
LDA	48.9	18.3	76.6
QDA	48.3	52.3	51.3
LOGDA	49.1	19.0	75.6
KNN1	46.0	46.4	59.9
Soft computing methods			
ANFIS	22.0	64.0	84.5
NNET	20.0	76.1	83.3

First we carefully selected the parameters by comparing the different definitions in literature and by clearly stating any modification introduced in the numerical procedures. FHR signal quality assessment was considered essential [12]. Numerical indexes were computed on short 3 minutes windows and averaged to reduce intraindividual variability.

Second we tested on this set of parameters different methodological approaches to the discrimination of pathological cases. Classical supervised classifiers fail to distinguish pathological from normal fetuses. It may be possible that the normality hypothesis, required by quadratic discriminant analysis (DA) and logistic DA, is not appropriate for a few variables included in the parameter set. The poor value of the true classification rates obtained also with linear DA, probably suggest that the two populations lie in very convoluted and intermingled regions in the parameters space. Direct inspection of the data set confirmed such assumptions. Therefore, only methods able to shape very complex decision regions are eligible to succeed.

The ANFIS and MLP algorithms achieved both about 80% true classifications rate with sufficient high sensitivity and specificity. We acknowledge that the methods need to be checked with a large database of CTG recordings before they can be used in the clinical environment, but they have been setup with a larger clinical study than any other similar approach [13, 14]. Moreover the results are encouraging and they were achieved by a completely automatic procedure.

Very preliminary results, obtained by the combination of both techniques with a simple rule inference system dealing with the gestational age of the fetuses, seem to be promising. This solution is only one among the possible future improvements.

REFERENCES

[1] H.P. Van-Geijn, “Developments in CTG analysis,” *Baillieres Clin. Obstet. Gynaecol*., vol.10(2), pp.185-209, 1996

[2] K.J. Dalton and A.J. Dawson, “Baseline: A computer method of calculating baseline in fetal heart rate recordings”, *Int J Biomed Comput*, 15, pp.311-317, 1984.

[3] R. Mantel, H.P. van Geijn, F.J.M. Caron. et al.,”Computer analysis of antepartum fetal heart rate”, *Int J Biomed Comput*, 25, pp.261-272, 1990

[4] W.J. Parer, J.T. Parer, R.H. Holbrook, B.S.B. Block, “Validity of mathematical methods of quantitating fetal heart rate variability”, *Am. J. Obstet.Gynecol*., Vol. 153, n.4, pp. 402-409, 1985.

[5] D. Arduini, G. Rizzo, A. Piana, et al., “ Computerized analysis of fetal heart rate:I description of the system (2CTG)”, *J Matern Fetal Invest* , 3, pp. 159-163, 1995

[6] G.S. Dawes, M. Moulden, C.W. Redman, “System 8000:Computerized antenatal FHR analysis”, *J Perinat Med.* , 19, pp 47-51, 1991

[7] S.M. Pincus, “Approximate entropy (ApEn) as a complexity measure”, *Chaos*, vol. 5, pp.110-117, 1995.

[8] G. Magenes, M.G. Signorini, D. Arduini, “Classification of cardiotochographic record by neural networks,” *Proc. IJCNN 2000, ISBN 0-7695-0619-4*, IEEE press, 2000.

[9] R. Sassi, “Analysis of Heart Rate Variability Complexity through Fractal and Multivariate Approaches”, PhD Thesis, Politecnic of Milan, 2000.

[10] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS*, 3_{rd} ed. Springer, New York, 1999.

[11] B. D. Ripley, “Classification (update)”, in *Encyclopedia of Statistical Sciences*, vol. Update 1, S. Kotz, C. B. Read and D. L. Banks Eds., John Wiley & Sons, New York, pp. 110–116, 1997.

[12] Task Force of the Europ. Soc. of Cardiol. & North Am. Soc.of Pacing and Electrophys., “Heart Rate Variability, standard of measurement, physiological interpretation and clinical use,” *Circulation*, vol.93:1043-1065, 1996.

[13] S. Kol, I. Thaler, N. Paz and O. Shmueli, “Interpretation of nonstress tests by an artificial neural network,” *Am. J.l of Obst. and Gynaecol*, vol. 12, pp.1372-1379,1995.

[14] K Maeda., M.Utsu, A. Makio, M. Serizawa, Y. Noguchi, T. Hamada., K. Mariko and F. Matsumoto, “Neural Network Computer Analysis of Fetal Heart Rate,” *J. Matern. Fetal Invest.*, vol.8, pp.163-171, 1998.